

Acht Aufgaben zur Vorbereitung Ihrer Daten für die Veröffentlichung

Aufgabe 1: Erstellen Sie Ihren eigenen Datenmanagementplan

Verwenden Sie die folgende Vorlage, um Ihren eigenen Datenmanagementplan für eines Ihrer Projekte zu skizzieren.

DMP-Vorlage¹

1. Allgemeine Angaben

- Details zum Projekt
 - Projektname
 - Beteiligte Personen/ Autoren
 - Institution
 - Anlass (Promotion, Drittmittelprojekt etc.)
 - Zeitraum
 - Kurzfassung

2. Datenerhebung und methodischer Hintergrund

- Forschungsdesign und Art der resultierenden Daten
- Zu berücksichtigende Fragen:
 - Werden die Daten generiert oder wiederverwendet?
 - Welche Datentypen werden in welchen Formaten erstellt/verarbeitet?
 - Wie groß ist das Datenvolumen?
 - Welche Geräte (Instrumente, Hardware, Software etc.) werden verwendet?
 - Wie sind/werden die Daten organisiert?
 - Wie werden der Forschungsprozess und die Daten dokumentiert?

3. Speicherung, Backup, Sicherheit

- Backup-Strategie und ggf. Umgang mit sensiblen Daten
- Zu berücksichtigende Fragen:
 - Wo werden welche Daten gespeichert?
 - Welche Speicherkapazitäten werden benötigt?
 - In welchen Abständen werden die Daten gesichert?
 - Sind Schutzmaßnahmen für sensible Daten erforderlich?
 - Müssen Dritte während der Projektlaufzeit auf Daten zugreifen?

4. Archivierung

- Maßnahmen zur Sicherstellung der langfristigen Verfügbarkeit der Daten
- Zu berücksichtigende Fragen:
 - Welche Daten sollen archiviert werden?

¹ Modifiziert von: Neumann, Janna, & Soßna, Volker. (2018, 22. März). Forschungsdaten managen - Anforderungen, Methoden, Hilfsmittel. Zenodo. <https://doi.org/10.5281/zenodo.1205287>, Folien 37-43, lizenziert unter [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

- Wo sollen die Daten archiviert werden?
- Wie lange sollen die Daten archiviert werden?
- Müssen für die Datenarchivierung rechtliche Fragen geklärt werden?
- Welche Kosten fallen für welche Leistung an?

5. Daten teilen und veröffentlichen

- Datenveröffentlichung und andere Maßnahmen, um Daten Dritten zugänglich zu machen
- Relevante Fragen:
 - Werden die Daten während der Projektlaufzeit an Dritte weitergegeben?
 - Wo sollen die Daten veröffentlicht werden?
 - Welche Metadaten müssen bereitgestellt werden, damit die veröffentlichten Daten auffindbar sind?
 - Welche zusätzlichen Informationen werden benötigt, um den Kontext der Daten zu verstehen?
 - Welche Lizenz soll vergeben werden?
 - Wird es Zugangsbeschränkungen geben?

6. Ressourcen und Verantwortlichkeiten

- Aufteilung der Aufgaben und Verantwortlichkeiten innerhalb des Projekts
- Benötigte Ressourcen aller Art (Kosten, Personal, Material)
- Relevante Fragen:
 - Wer ist für das Datenmanagement verantwortlich?
 - Welche personellen Ressourcen werden für eine erfolgreiche Umsetzung des Datenmanagementplans benötigt?
 - Wie hoch sind die Kosten für die Teilbereiche zur Umsetzung des Datenmanagementplans?
 - Welche zusätzlichen Infrastrukturressourcen werden benötigt und welche Kosten fallen an?

Aufgabe 2: Organisation Ihrer Forschungsdaten

Um Ihre Dateien zu strukturieren und ihnen aussagekräftige Namen zu geben, kann es hilfreich sein, sie physisch zu machen. Die folgende Übung zeigt Ihnen, wie das geht:

1. Schreiben Sie jeden vorhandenen Datei- und Ordernamen auf ein Post-it. Dies kann in physischer Form erfolgen, wenn Sie Post-its zu Hause haben, oder digital mit Microsoft 'Sticky Notes' oder einem entsprechenden Programm für Mac (wie Memo). Wenn Sie einen großen Datensatz haben, konzentrieren Sie sich nur auf einen Teil, z.B. einen Ordner, der sehr unordentlich ist.
2. Ordnen Sie die Ordner in einer sinnvollen Hierarchie an.
3. Legen Sie Ihre Dateien in die strukturierten Ordner.
4. Benennen Sie Ihre Dateien und Ordner auf vernünftige Weise um. Verwenden Sie eine vordefinierte Namenskonvention wie 'Datum_Kurzbeschreibung_Autor_Version'.
5. Benötigen Sie neue Ordner? Haben Sie zu viele?

Weitere Informationen zur Dateibenennung und -strukturierung finden Sie hier:

<https://www.forschungsdaten-bildung.de/dateien-bennen>

https://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf

Aufgabe 3: Wählen Sie Ihre Daten zur Archivierung oder Publikation aus²

Teilen Sie die von Ihnen generierten Forschungsdaten in die folgenden Datenklassen ein:

1) Sehr wertvolle Daten

Forschungsdaten oder Codes, die

- einzigartig oder sehr kostspielig zu restaurieren und
- für eine große Community und/oder für vielfältige Nachnutzungen interessant sind

➔ Diese Forschungsdaten sollten unbedingt archiviert und/oder publiziert werden!

2) Begrenzte wertvolle Daten

Forschungsdaten oder Codes, die entweder

- mit vertretbarem Aufwand wiederhergestellt werden können oder
- nur für kurze Zeit, für eine kleine Zielgruppe oder nur zum Nachweis von Forschungsergebnissen interessant sind

➔ Diese Forschungsdaten können archiviert oder publiziert werden. Es ist jedoch nicht erforderlich, wenn Aufwand oder Kosten dagegen sprechen.

3) Wenig wertvolle Daten

Forschungsdaten oder Codes, die

- nur zu Testzwecken erstellt werden, oder
- nicht lesbar oder nicht ausreichend geschrieben wurden

➔ Diese Forschungsdaten müssen nicht archiviert oder publiziert werden.

² Modifiziert von: TU9-FDM. (2020, 10. März). Umgang mit großen Datenmengen. Zenodo. <https://doi.org/10.5281/zenodo.3702113>, Folie 28, lizenziert unter [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)

Aufgabe 4: Geeignete Dateiformate für die Langzeitarchivierung und Publikation finden

Die Dateien Ihrer Forschungsdaten sollten in standardisierten, weit verbreiteten, offenen Dateiformaten für nicht-proprietäre Programme verfügbar sein.

Eine Auswahl geeigneter Dateiformate ist in der folgenden Tabelle aufgeführt³:

	Empfohlene Formate	Weniger empfehlenswert	Konvertierungsmöglichkeiten
Textformate	PDF/A, TXT (UTF8-kodiert), XML	DOC, PPT	Word & PowerPoint zu PDF/A LaTeX & TeX zu PDF/A
Tabellen/ Tabellenkalkulationen	CSV/TVS	XLS	Tabellen zu ASCII-Textdateien (CVS)
Statistische Umgebungen	SDD-Dateien, R-Dateien, Matlab-Dateien im HDF-Format		Speichern Sie Matlab-Dateien als v7.3 MAT (HDF5-basierter Standard) Speichern Sie wichtige Tabellen im Arbeitsbereich separat als CSV
Rastergrafiken	TIFF, PNG, JPEG2000	EPS	
Vektorgrafiken	SVG		
Multimedia	WAV, AVI, MOV, MPEG-4	WMV	
Geodaten	GML, KML, ESRI Shapefile, georeferenziertes TIFF		
Numerische Daten	NetCDF		

Weitere Informationen finden Sie hier:

<https://www.ianus-fdz.de/it-empfehlungen/dateiformate>

<https://www.loc.gov/preservation/digital/formats/fdd/descriptions.shtml>

<http://www.nationalarchives.gov.uk/PRONOM/Default.aspx>

Prüfen Sie, ob Ihre Daten bereits ein geeignetes Format haben.

Falls nicht, prüfen Sie, ob eine Konvertierung in ein geeignetes Format möglich ist. Es sollte darauf geachtet werden, dass weder der Zeitaufwand noch ein möglicher Datenverlust zu groß werden.

Es gibt mehrere Möglichkeiten und Werkzeuge für die Konvertierung. Kann Ihr Datenformat problemlos konvertiert werden? Wenn nicht, welche anderen Lösungen könnten Sie finden?

³ Modifiziert von: ETH Zürich Bibliothek Fachstelle Digitaler Erhalt, Archivtaugliche Dateiformate, <http://www.library.ethz.ch/ms/Digitaler-Datenerhalt-an-der-ETH-Zuerich/Downloads>, zuletzt aufgerufen August 2021

Aufgabe 5: Beschreiben Sie Ihre Daten mit geeigneten Metadaten

Welche Informationen wären grundsätzlich notwendig, damit ein Außenstehender Ihre Daten verstehen und weiterverwenden kann?

Diese Frage können nur Sie für Ihre Forschungsdaten beantworten. Es kann jedoch hilfreich sein, fachspezifische Metadatenschemata zu verwenden. Dies sind Sammlungen von Elementen, die verwendet werden, um Forschungsdaten zu beschreiben.

Fachspezifische Metadatenschemata finden Sie hier:

<https://www.dcc.ac.uk/resources/metadata-standards>
<https://rd-alliance.github.io/metadata-directory/>

Sie können gerne eine README-Vorlage verwenden. Hier sind einige Beispiele:

<https://cornell.app.box.com/v/ReadmeTemplate>
<https://datamanagement.hms.harvard.edu/collect/readme-files>

Überlegen Sie, ob die benötigten Informationen ausreichen, um Ihren Datensatz zu verstehen.

Überlegen Sie beispielsweise, welche Vorarbeiten, Korrekturen und Berechnungen Ihre Daten durchlaufen haben, wann und woher die Daten stammen oder mit welcher Software sie generiert wurden. Bei Tabellen ist es wichtig zu erklären, was Zeilen- und Spaltennamen ausdrücken.

Aufgabe 6: Alle rechtlichen Aspekte berücksichtigt?

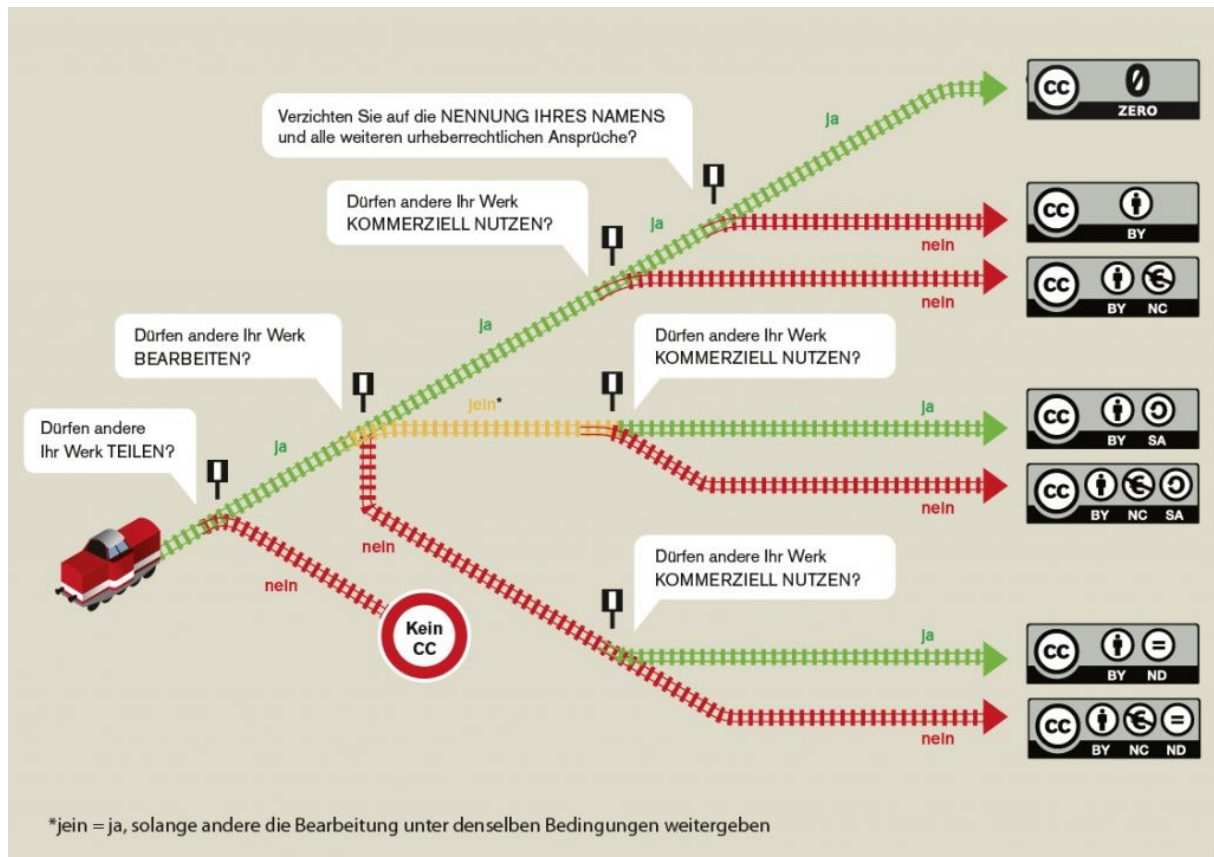
Dieses Flussdiagramm⁴ hilft Ihnen zu prüfen, ob die wichtigsten rechtlichen Aspekte bei der Veröffentlichung Ihrer Forschungsdaten berücksichtigt wurden.



⁴ Modifiziert von „[Hindernisse beim Veröffentlichen von Forschungsdaten](#)“ von Peter Brettschneider lizenziert unter [CC BY 4.0](#)

Aufgabe 7: Wählen Sie Ihre Lizenz!

Um eine passende Lizenz für Ihre Forschungsdaten zu finden, beantworten Sie die folgenden Fragen der Infografik⁵ und finden Sie heraus, welche Lizenz am besten geeignet ist.



Diese Auswahlmethode gilt nur für allgemeine Forschungsdaten. Wenn Sie Forschungssoftware oder eine Datenbank veröffentlichen möchten, können Sie weitere Informationen über folgende Links erhalten:

<https://ifross.github.io/ifrOSS/Lizenzcenter>

<https://choosealicense.com/>

⁵ „Welches ist die richtige CC-Lizenz für mich?“ von Barbara Klute und Jöran Muuß-Merholz für [wb-web](#) lizenziert unter [CC BY SA 3.0](#)

Aufgabe 8: Finden Sie ein geeignetes Repository

Wie finde ich ein geeignetes Repository?

- Institutionelles Repository an der eigenen Forschungseinrichtung: Die TU Braunschweig bietet die [Institutionelles Repository](#)
- Fragen Sie Kollegen (auch aus Kooperationen)
- Suchen Sie auf re3data.org - Registry of Research Data Repositories